

Section 6. Solutions

This section describes the process for the identification and comparison (pros and cons) of approaches and technology solutions that will contribute to the EarthCube goal of satisfying current and future research needs of the geoscience end-user.

6.1 The Problem Domain

There are many earth science simulation use cases that appear to leverage different approaches in their inherent workflows. As noted in section 1 (Purpose), discussions regarding this phenomenon of different modes of use of workflows across the user community led to an initial formulation of “Workflow Use Paradigms”. The initial list of identified paradigms were described in Section 1.3.1.

The challenge is that, despite the many inherent workflows that are present in geoscience research areas, a significant proportion of the scientists do not leverage technology mechanisms to automate the discernible workflows in their research work. The underlying issue of lack of workflow automation appears to be more a case of ‘lack of motivation towards technology adoption’ or ‘difficulties in technology adoption’ rather than being an issue of inability to determine the existence of workflows in the geoscientists’ work. This conclusion is discussed in section 3 (challenges) and is supported by the responses to a survey conducted by the workflow community group as detailed below:

In response to the question; “Improving the efficiency of a complicated or error-prone set of computational steps or tasks is a common problem in research. Do you have these types of problems in your research? How do you address them?”, none of the respondents claimed that they *do not* have an ‘error prone set of computational steps or tasks’. Yet their feedback on how they addressed them varied greatly with responses including strategies such as; ignoring, writing down the steps in a notebook or wiki, detailed documentation, writing shell scripts, and creating reusable workflows. Yet only a couple of respondents followed with feedback for the related question of, “what are the strengths and weaknesses of your current approach? What improvements would you like to see?”. Thus, the motivation to adopt technology to solve their problems appears to be low.

These findings indicate that geoscientists need to be guided through the process of discerning the inherent workflows in their research work, and the subsequent automation of those workflows. The survey results also indicate that the approaches towards use of automated workflows in geoscience research would be very diverse and dependent on factors such as,

- the workflow use paradigm
- the current method of dealing with the inherent workflow

- the nature and characteristics of the underlying problem domain
- the stages at which the workflow is apparent in the research process - data collection stage, computational stage, data analysis etc.
- the technical expertise available to the geoscientist or the geoscience research team
- the level of access to cyberinfrastructure

Further, the appropriate approach and suitable technology solutions would be dependent on the underlying objectives of the geoscience research team, the larger geoscience domain or a large organization covering many geoscience areas. Such objectives may include:

- The increased use of existing workflow tools by geoscientists.
- The improvement of existing tools or development of new tools to cover gaps identified by the Status and Requirements Task Force.
- The targeted use of workflow technologies in very high profile projects that aren't currently using them.
- The increased ability to document and reproduce research results and community data products using workflows.
- The extension of workflow technologies or development of new ones to solve grand challenge problems.
- The sustainability of tools.

Many of these factors and objectives are noted in the previous sections of this roadmap document. The processes discussed in this section will be primarily executed by the Engagement Task Force.

The Workflows Working Group will include an Engagement Task Force that will: 1) provide guidance to geoscientists in identifying a approaches to address their workflow needs, 2) assist scientists in evaluating potential workflow technology solutions, 3) request the support of the Status and Requirements Task Force and the Prototyping Task Force when necessary 4) disseminate expertise in workflow solution approaches.

6.2 Processes to Identify Approaches

In many domains beyond geoscience, maturity models have been leveraged as guidance mechanisms to systematically achieve increasing levels of sophistication in desirable capabilities and characteristics. Some domains such as software development adopt a single maturity model such as CMMi across all organizations. Yet some domains such as eGovernment and eHealth tend to specify organization-specific maturity models, borrowing heavily from existing models and customizing them to suit their purposes.

Considering the diverse nature of geosciences and the multifaceted factors listed above, the process of identifying an approach to that would systematically introduce workflow orchestration

and automation capabilities in a geoscience research group could be guided by the definition of an organization-specific “Workflow Capability Maturity Model”. Such a workflow capability maturity model will define levels of maturity that could provide the aims and directions of growth in workflow orchestration and automation sophistication within the organization. Each level would detail the milestones and metrics to measure progress. In this manner a workflow capability maturity model will serve as a tool to reflect upon the organization’s progress in effectively utilizing workflows to address their problem domain specific orchestration and automation goals.

Support for the creation of such a maturity model for each organization would be provided by Engagement Task Force, to whom the the “task” would be handed off, after the work of the Status and Requirements Task Force.

The Engagement Task Force would be able to help the target organization identify their path of maturing capabilities in workflow orchestration and automation by considering the evident ‘Workflow Use Paradigms’, their unique domain factors and needs, and any specific requirements that may necessitate the involvement of other community groups such as data mining, brokering and interoperability.

6.3 Processes to Identify Technology Solutions

Once the potential involvement of other community groups such as data mining, brokering and interoperability have been established and an appropriate approach to handle the problem domain has been structured through the design of a workflow capability maturity model, the Engagement Task Force will be involved in helping the geoscience team identify technology solutions.

As noted in section 5 (Status), a plethora of wide and varied technology solutions exist, and the process of standardizing and refining these diverse systems, if agreed upon, would be an ongoing and long term process. Thus identifying technology solutions that are aligned with the solution approach to the specific problem domain would be a tricky and non-trivial exercise with the long-term impact. In addition to working with the other community groups that would be associated in this effort, the requisite standards that should be adopted will also need to be determined.

Therefore, the process to identify technology solutions should be done in a systematic manner through the definition of a “Technology Evaluation Framework”. Such a framework should be designed for each technology solution related goals or tasks that evolve from the progress through the workflow capability maturity model. A technology evaluation framework would specify the comparison characteristics and associated metrics, including absolute conditions that need to be satisfied by a particular technology solution as well as specifics with respect to the process of evaluation such as requisite testbeds, infrastructure, prototyping needs etc.

The Engagement Task Force will provide the necessary guidance for this, and when necessary, will obtain the support and assistance of other task forces. For example, the Prototyping Task force would most likely be needed in the cases of multiple technology solution evaluation or in the cases of complex technology solution integration.

There will be a need to assess the adoption and impact of workflow technologies as a result of EarthCube activities. This will require defining metrics to measure adoption and impact, and to collect data as EarthCube progresses. In the early stages of the project, baseline data should be collected to assess the state of the art and the initial levels of adoption. As the activities of the Workflows Working Group progress, additional data should be collected to assess adoption and impact. The data should be analyzed, particularly to detect barriers and challenges that will inform and adjust the roadmap activities.

The Assessment Task Force will track and assess the impact of workflow technologies across geosciences through: 1) defining metrics to measure impact in geosciences, 2) collecting quantitative and qualitative data at the early stages of EarthCube as baselines, 3) collecting additional data to measure and demonstrate progress as the roadmap activities progress, 4) analyzing the collected data to understand any issues and challenges that may need to be resolved in order to achieve the roadmap goals.